

## Editorial

# Determining the sample size necessary to pass the tentative final monograph pre-operative skin preparation study requirements

Daryl S. Paulson\*

BioScience Laboratories, Inc. 1765 South 19th Avenue, Bozeman, Montana 59718, USA

### \*Correspondence:

Dr. Daryl S. Paulson,

E-mail: [dpaulson@biosciencelabs.com](mailto:dpaulson@biosciencelabs.com)

**Copyright:** © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

How can one predict the sample size needed to pass the current FDA's requirements for the preoperative skin preparation with two recent additions (the 95% confidence intervals at 2.0 and 3.0 log<sub>10</sub> reductions and 70% responder rates)? The purpose of the sample size formula presented by the Food and Drug Administration (FDA) is able to detect if the test and positive control products are the same or different. It can also tell if the test product achieves greater reductions than the vehicle or the negative control product. The sample size formula for this is:

$$n \geq \frac{x s^2 \left( z_{\alpha/2} + z_{\beta} \right)^2}{d^2}$$

where:

$n$  ≥ sample size needed per product arm of the study

$x$  = the number of different products in the study (usually products but if a vehicle or negative control is used, then three products; sometimes the FDA requests more)

$s^2$  = largest variance of the product samples at times 30 seconds or 10 minutes

$z_{\alpha/2}$  = 1.96 = Type I error ( $\alpha$ ) = 0.05 (Confidence level = 0.95 = 1 - 0.05)

$z_{\beta/2}$  = 0.842 = Type II error ( $\beta$ ) = 0.20 (Power = 0.80 = 1 - 0.20)

$d$  = Detectable difference = 0.5 log<sub>10</sub>. This is the value that the FDA currently uses.

The FDA uses a medical or biostatistical approach, meaning both types of errors ( $\alpha$  and  $\beta$ ) are important in the analysis. Again, the purpose of this sample size formula is to provide the number of subjects required for each test and positive control product (and perhaps the vehicle [negative control] product), so 1) one can conclude there is a difference when, in fact, the products are different 95% of the time, and 2) that the products are

the same when, in fact, they are the same 80% of the time.\* If there is no difference between the test product and the positive control product, the FDA's 1994 tentative final monograph (TFM) states this finding be in the study's final report.<sup>1</sup> This is how the TFM was originally written with the sample size formula.

However, the FDA later included two additional requirements of log<sub>10</sub> reductions and 70% responder rates with the 95% confidence intervals in these two additions. Neither the 1994 nor the 2015 version of the Tentative Final Monograph even discuss these two components from a statistical perspective. Yet they are there, making it more difficult to pass the TFM requirements.

Specifically, the new requirement states that the lower bounds of the 95% confidence interval must be greater than or equal to 3.0 log<sub>10</sub> reductions for the inguinal site and greater than or equal to 2.0 log<sub>10</sub> reductions for the abdominal site for both the test and positive control products at the first sample time (30 seconds or 10 minutes). In addition, the 70% responder rates must have the lower bounds of the 95% confidence interval greater than or equal to 70% for both the test and positive control products at the first sample time (30 seconds or 10 minutes).

Unfortunately, the original sample size calculation formula does not address these last two requirements. What can one do to pass the evaluation, especially with the 95% confidence interval of the 70% responder rate?

\* This is more complicated than described, because there are two sides (left and right) on each subject. If there were only two products used, then no more subjects than the  $n$  value will be needed. If there are three products, then there will be more subjects used.

**Pilot study**

A company first should determine if their test product and the positive control product can achieve these extra requirements. If they can, then how many subjects must be evaluated to reduce the 95% confidence intervals to be greater than or equal to 2.0 or 3.0 log<sub>10</sub> reduction and be within the 70% responder rates? (See appendix)

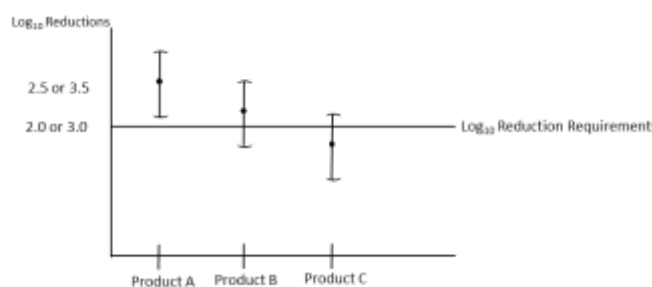
One may question if a pilot study is necessary when one “knows” one’s product will pass, based on other work done. The new position is that the laboratory doing the work must calculate the number of subjects required to pass the 2.0 and 3.0 log<sub>10</sub> reductions, but especially the 70% responder rate with the lower bounds of the 95% confidence interval greater than or equal to 70%. There is no way to know without performing a pilot study.

Conducting a pilot study can spare the headache of discovering the product does not pass upon the conclusion of a large pivotal or confirmatory clinical trial. If the product does not pass the pilot study, there are four areas to examine:

- type of applicator, including sponge type or other material(s) used,
- the amount of antimicrobial product available within the applicator,
- the formulation (alcohol + CHG, alcohol + PVP-I, etc.), and
- developing a suitable method of product application.

**What the statistics should do**

The same model – an analysis of variance (ANOVA) in the pilot study as in a pivotal or confirmatory study – should be used to determine if the products will pass. For example, let us examine three scenarios, looking at only the test product for simplicity. The log<sub>10</sub> reductions are illustrated in Figure 1 below:



**Figure 1: Log<sub>10</sub> reductions at abdominal and inguinal sites.**

**Scenario A**

In instances like this, Product A passes the log<sub>10</sub> reduction with a 2.5 log<sub>10</sub> (abdomen) or 3.5 log<sub>10</sub> (inguinal), then one should also compute the

70% responder rate and see if the products pass with the proportion of successes above 75%. If they do, then calculate the total number of subjects needed to pass the evaluation with the 95% confidence limits equal to or greater than 70%.

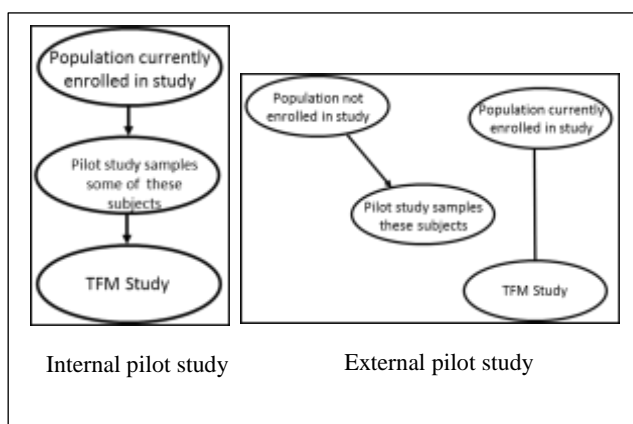
**Scenario B**

In this case, it will be necessary to use more subjects to achieve a 95% confidence interval of Product B to pass the test. However, passing the 70% responder rate is questionable. One needs to achieve higher reductions (2.5 or 3.5 log<sub>10</sub>) with the product. Review the four areas just discussed, and see what could be changed to achieve more reductions.

**Scenario C**

Product C will not pass, no matter how many subjects are used, because the mean is below the required log<sub>10</sub> reduction (see previous section about addressing the four areas of product development).

Formerly, companies chose to do a pilot study at the facility performing the pivotal study to calculate the sample size needed to pass both studies. All too often, they relied on the previously-stated FDA sample size formula, instead of basing it on the two additional requirements. Also, because there is a difference in the populations of subjects between the two laboratories who perform the pivotal and confirmatory studies (climate, biological, ethnicity, and the laboratory methods, personnel, and standard operating procedures [SOPs]), this no longer works, because the criteria are much stricter. The pilot study should also use a greater number of subjects (15 – 20) to derive more precise estimates than the previous pilot studies (5 – 10).



**Figure 2: Internal vs. external pilot study.**

It would be better to conduct an internal pilot study.<sup>2</sup> An internal pilot study uses a random sample of subjects from the population of subjects who have already been enrolled in this study. An external pilot study uses subjects who are not currently enrolled in the study. The internal and external studies are depicted in Figure 2.

The internal pilot study is more accurate and precise in discovering if the products will pass the TFM requirements. Because many of the calculations would be drawn from the same subjects, the FDA has stated that using this method is not acceptable. So we must rely on an external pilot study.

**70% Responder rates**

Originally, the FDA wanted to make certain the distribution was normal without skewed data or outliers to the right. If outliers were present in the data, the mean would shift to the right, indicating higher reductions. Yet, this could be better done with exploratory data analysis (EDA), using stem-and-leaf displays, *R*-Values, and Boxplots.<sup>3</sup>

The original formula for this computation was:

$$\pi = p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} \pm \frac{1}{2n}}$$

Where

$\pi$  = 95% confidence interval of responder rate success

$p$  = proportion of success  $\frac{\text{number of successes}}{\text{number of total observations}}$  †

$n$  = sample size of this portion of the study (one product and one sample time)

$z_{\alpha/2}$  = 95% confidence interval =  $\frac{0.05}{2} = 0.025$ .

$z = 1.96$  for a two-tail confidence interval.

The Yates correction factor was also used, which is  $\frac{1}{2n}$

or  $\frac{0.5}{n}$ .

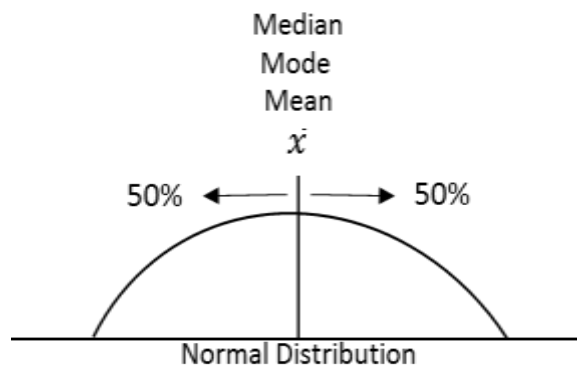
However, the Yates correction factor is no longer necessary, according to the FDA's comments on the latest studies. So, the formula now is

$$\pi = p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

This makes the 95% confidence interval narrower and easier to pass. However, to pass the 70% responder rates with the lower bounds of the 95% confidence interval at this level is quite difficult.<sup>4</sup>

Nevertheless, in order to pass, the  $\log_{10}$  reductions must be equal to or greater than 2.5 and 3.5, respectively for the abdomen and the inguinal sites.

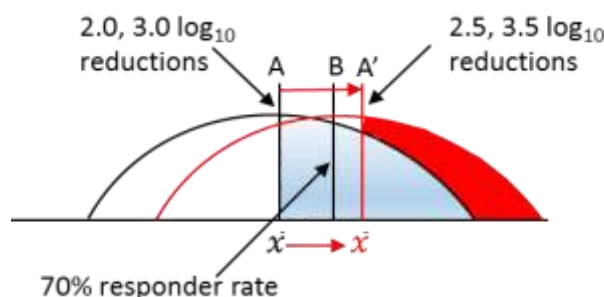
Looking at this situation in greater detail, for a normal Gaussian distribution, the median, mode, and the mean have equal percentages of 50% data below and 50% data above the mean as in Figure 3.



**Figure 3: Normal distribution.**

If a product achieved a 2.0 or 3.0  $\log_{10}$  reduction, according to Figure 3, this would result in a responder rate of 50%. But the FDA raised the requirement to 70%. The only way to pass is to have much higher reductions than 2.0 or 3.0  $\log_{10}$ . Figure 4 demonstrates the necessary adjustment. Line A represents the current mean (which is 2.0 or 3.0  $\log_{10}$  reductions), which is to the left of the Line B that represents the 70% responder rate. And, of course, the product will not pass.<sup>‡</sup>

The product must perform better, and to do this will require the mean ( $\bar{x}$ ) of the data to shift to the right, as demonstrated by Line A', to a point beyond the 70% responder rate ( $\geq 75\%$ ) (Line B). To ensure the product passes, the mean should fall at or above the 2.5 or 3.5  $\log_{10}$  reductions, illustrated by Line A'. Now the lower limits of the 95% confidence interval must be greater than the 70% responder rate by a sample size calculation. It is done by setting the lower limit to 0.70 and solving for  $n$ .<sup>5</sup>



**Figure 4:  $\log_{10}$  reduction.**

† “0” = lower than the 2.0 (for the abdomen) or 3.0 (for the inguinal)  $\log_{10}$  reductions; “1” = equal to or higher than the 2.0 (for the abdomen) or 3.0 (for the inguinal)  $\log_{10}$  reductions. The values that received the “1s” are successes, and the “0s” are failures.

‡ I have not accounted for the lower bounds of the 95% confidence interval in order to keep the discussion simple.

In summary, this requires two points are met:

- the percent responder rate must be  $\geq 0.75$  (75%), and
- increase the number of subjects used to tighten up the 95% confidence interval of the responder rate.

However, these rules are not absolute. If, for example, your product gets a responder rate of 74%, this will be alright but will require more subjects than if it was at 75%.

## CONCLUSION

It is important not to be overly concerned with the FDA's sample size formula listed in the TFM. However, the number of subjects used should be equal to or greater than the TFM proposed guidelines. It is far more important to conduct a pilot study at the laboratories selected to run the studies 1) to see if the test and control products will pass the 70% responder rate, and 2) to then determine the sample size correctly to reduce the 95% confidence level to be 70% or above.

## REFERENCES

1. Food and Drug Administration's Tentative Final Monograph, 1994. Available at <http://www.fda.gov/OHRMS/DOCKETS/98fr/94-14503.pdf>. Accessed on 10 July 2016.
2. Ryan TP. *Sample Size Determination and Power*. Hoboken, NJ: John Wiley & Sons; 2013: 20 – 24.
3. Velleman PF, Hoaglin DC. *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston, MA: Duxbury Press; 1981.
4. Paulson DS. *Topical Antimicrobial Testing and Evaluation*. 2nd Edition. Boca Raton, FL: CRC Press; 2015: 102 – 104.
5. PASS Version 14. Available at <https://www.statcon.de/shop/en/software/statistics/pass>. Accessed on 10 September 2016.

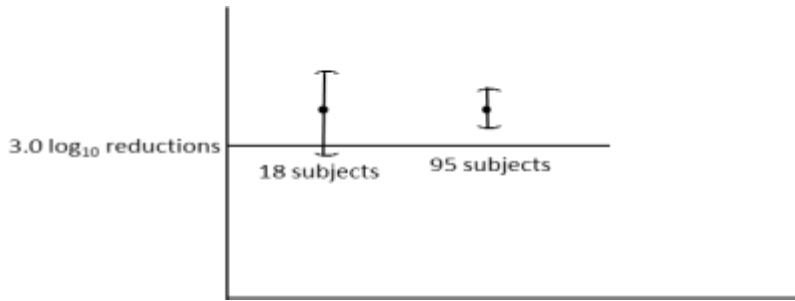
**Cite this article as:** Paulson DS. Determining the sample size necessary to pass the tentative final monograph pre-operative skin preparation study requirements. *Int J Clin Trials* 2016;3(4):169-73.

## APPENDIX.

As  $n$  gets larger, the 95% confidence interval gets smaller. If  $s = 2.2$ , then  $\bar{x} = 3.5$ . How it works is shown on A and B, using 18 and 95 subjects.

	$\mu = \bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$				
A	$\frac{2.2}{\sqrt{18}} = 0.5185$	$n = 18$	$\mu = 3.5 \pm 1.96 \frac{2.2}{\sqrt{18}}$	$2.48 \leq \mu \leq 4.52$	This is wider. It will not pass the $3.0 \log_{10}$ reduction
B	$\frac{2.2}{\sqrt{95}} = 0.2257$	$n = 95$	$\mu = 3.5 \pm 1.96 \frac{2.2}{\sqrt{95}}$	$3.04 \leq \mu \leq 3.96$	This is tighter. It will pass the $3.0 \log_{10}$ reduction.

For example, see Figure 5.



**Figure 5: Confidence intervals.**

95% confidence interval =  $\mu = \bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$